



(11) **EP 0 751 464 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
02.01.1997 Bulletin 1997/01

(51) Int Cl.⁶ **G06F 13/40, G06F 15/16,
G06F 3/06**

(21) Application number: **96303626.4**

(22) Date of filing: **21.05.1996**

(84) Designated Contracting States:
DE FR GB

• **Chen, Shenze**
Cupertino, CA 95014 (US)

(30) Priority: **26.06.1995 US 494457**

(74) Representative: **Jehan, Robert et al**
Williams, Powell & Associates,
34 Tavistock Street
London WC2E 7PB (GB)

(72) Inventors:

• **Thapar, Manu**
Fremont, Ca 94555 (US)

(54) **Storage system**

(57) A multi-host storage system uses the dual port feature of a Fibre Channel to allow multiple computer hosts (312, 314) to simultaneously access a cluster of memory units (301, 315) that are Fibre Channel arbitrated devices provides for fault tolerance and redundancy. Typical multiple host-access schemes require an expensive Fibre Channel switch and do not allow simultaneous accessing. The dual port feature of Fibre Channel devices provides for fault tolerance and redundancy.

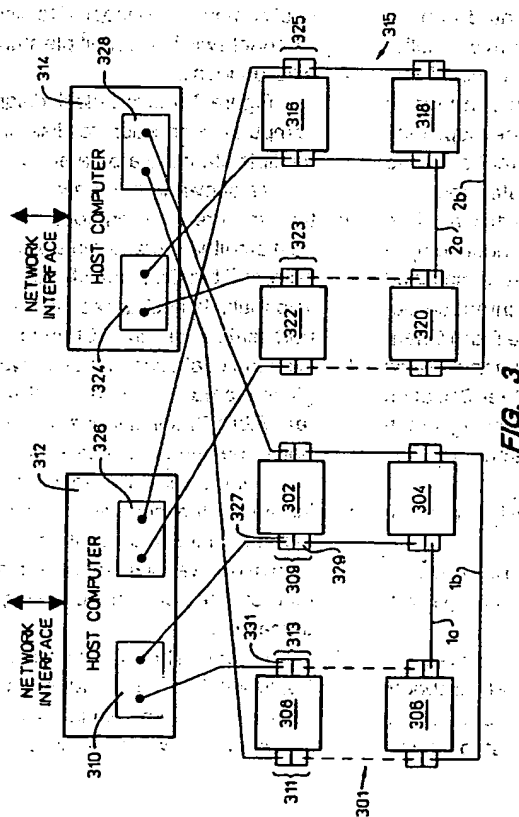


FIG. 3

EP 0 751 464 A1

Description

The invention pertains to a multi-host information storage system architecture Fibre Channel arbitrated loops for connecting disk drives.

Fibre Channel is the general name of an integrated set of standards being developed by the American National Standards Institute (ANSI) which defines new protocols for flexible information transfer. The Fibre Channel Standard addresses the need for very fast transfers of large volumes of information, while at the same time relieving system manufacturers from the burden of supporting the various types of channels and networks currently focusing on niche requirements.

Fibre Channel defines an arbitrated loop topology for connecting disk drives to host computers. Figure 1 shows a block diagram of the connections required between a host computer and a plurality of disk drives to implement the arbitrated loop topology. A host computer 102 has a Fibre Channel controller 104. The Fibre Channel controller 104 controls the disk drives 106, 108, 110, 112 by sending commands such as reads and writes to the drives. Each individual drive has a burst data access rate and a sustained data access rate. Burst rates can be as high as 100 MB/second. Sustained data access rates are much lower. By arbitrating between disk drives, the combined sustained data access rate of the plurality of drives can be as high as the peak rate of 100 MB/second.

For reliability and fault-tolerance, Fibre Channel defines a dual port configuration for devices attached to an arbitrated loop. The dual port configuration effectively allows two separate Fibre Channel arbitrated loops to be connected to the same cluster of disk drives. Reliability of the cluster of disks is enhanced because if one loop fails for some reason, the other loop can help to compensate for the failure. The two separate loops may be connected to one controller of a computer host, or to two separate controllers of a single computer host.

It is desirable to be able to access a particular Fibre Channel arbitrated loop by more than one host computer. Figure 2 shows how this is presently accomplished by using a Fibre Channel switch 201 to interface between the host computers 202, 204 and the Fibre Channel arbitrated loops 206, 208. The Fibre Channel switch and the added interfaces add costs. Another limitation of the Fibre Channel switch is that with a blocking switch only one cluster of Fibre Channel arbitrated loop disk drives may be accessed at a time.

The present invention seeks to provide improved information storage.

According to an aspect of the present invention, there is provided a multi-host information storage system architecture as specified in claim 1.

The invention can provide for multiple access to Fibre Channel devices to obtain greater shared bandwidth without adding an expensive switching interface. It is also possible for more than one computer host to be able

to access different disk drives of a Fibre Channel arbitrated loop disk cluster at the same time.

The preferred embodiment provides an apparatus and a method of multiple computer host access to Fibre Channel arbitrated loops that uses the dual port feature of Fibre Channel memory storage units to achieve higher aggregate bandwidth and sharing of data. This embodiment accomplishes these attributes without the overhead of an expensive Fibre Channel switch. The memory storage units of this embodiment are generally disk drives. This embodiment allows two different computer hosts to access different disk drives units of a given Fibre Channel arbitrated loop disk drive cluster at the same time.

This preferred embodiment includes a multi-host storage architecture comprising a plurality of hosts wherein each host has a plurality of controllers. Each one of the host controllers is connected to one of a plurality of Fibre Channel arbitrated loops. Each of the Fibre Channel loops comprises a plurality of memory storage units. Each of the memory storage units has a plurality of data transfer ports wherein each one of the data transfer ports of each memory storage unit is connected to a corresponding data transfer port of the adjacent memory storage unit within each loop.

An embodiment of the present invention is described below, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 shows a prior art block diagram of a host computer connected to disk drives through a standard Fibre Channel arbitrated loop.

Figure 2 shows a prior art block diagram of multiple host computers connected to disk drives through a Fibre Channel switch and multiple standard Fibre Channel arbitrated loops.

Figure 3 shows a block diagram of the multiple host computer connection to disk drives through preferred multiple standard arbitrated loops.

As shown in the drawings for purposes of illustration, the preferred embodiment is a multiple computer host architecture wherein each host has access to a plurality of Fibre Channel arbitrated loops of memory units. The multiple host access is obtained by using the dual port feature of Fibre Channel memory storage units. With the dual ports, either of two hosts may access any given memory unit contained within either of two separate Fibre Channel loops. By increasing the number of ports, the number of hosts that can access the memory unit will increase. Increasing the number of ports also increases the number of Fibre Channel loops that a memory unit can be contained within. The architecture also allows various ones of the multiple hosts to access different memory units of a single arbitrated loop at the same time.

Figure 3 shows a block diagram of a preferred embodiment. The system allows more than one computer host to access a particular Fibre Channel arbitrated loop of disk drives without a Fibre Channel switch.

As shown in Figure 3, a first cluster 301 of memory storage units is formed by a plurality of disk drives 302, 304, 306, 308. For this particular embodiment, the first cluster of memory storage units are connected to two Fibre Channel arbitrated loops. Fibre Channel arbitrated loop 1a is formed by connecting a first data transfer port 309 of each of the drives 302, 304, 306, 308 to a first controller 310 of a first host computer 312. The other Fibre Channel arbitrated loop 1b is formed by connecting a second data transfer port 311 of each of the drives 302, 304, 306, 308 to a second controller 328 of a second host computer 314. This configuration allows both the first host computer 312 and the second host computer 314 to access different drives within the first memory storage unit cluster 301 of the arbitrated loops 1a, 1b at the same time.

A second cluster 315 of memory storage units is formed by a second plurality of disk drives 316, 318, 320, 322. For this particular embodiment, the second cluster of memory storage units are connected to two Fibre Channel arbitrated loops. Fibre Channel arbitrated loop 2a is formed by connecting a first data transfer port 323 of each of the drives 316, 318, 320, 322 to a second controller 326 of the first host computer 312. The other Fibre Channel arbitrated loop 2b is formed by connecting a second data transfer port 325 of each of the drives 316, 318, 320, 322 to a first controller 324 of the second host computer 314. This configuration allows both the first host computer 312 and the second host computer 314 to access different drives within the second memory storage unit cluster 315 of the arbitrated loops 2a, 2b at the same time.

Figure 3 shows four disk drives in the first cluster 301 and the second cluster 315. However, the number of drives in the clusters 301, 315 can be increased.

The connections between the memory storage units and the controllers can be optical or electrical. If the connections are optical, then typically an optical fiber is used. If the connections are electrical, then typically a copper based cable is used. Optical fiber allows greater physical distance between connections, but is more expensive than copper. Fibre Channel specifications define the dimensions and characteristics of the cabling or fiber to be used.

The cable or fiber used is connected to the memory storage units through the data transfer ports 309, 311, 323, 325. The data transfer ports 309, 311, 323, 325 each have both an input and an output. Figure 3 shows the input 327 and the output 329 of the first data transfer port 309 of the memory storage unit 302. A cluster of memory units are Fibre Channel arbitrated loop configured by connecting the data transfer port output of each memory unit to the data transfer port input of a neighboring memory unit to form a loop. The two memory storage units at each end of the loop are configured so that the input portion of the data transfer port of one of the memory storage units (designated the input memory storage unit) is connected to an output of a controller

and the output portion of the data transfer port of the other memory storage unit (designated the output memory storage unit) is connected to the input of the same controller. For example, as shown in Figure 3, the two memory units at the end of loop 1a are disk drives 302, 308. The output of controller 310 is connected to the input 327 of the data transfer port 309 of the input memory storage unit 302. The input of controller 310 is connected to the output 331 of the data transfer port 313 of the output memory storage unit 308. The specifics of these connections are defined by the Fibre Channel specifications. The inputs and outputs of the data transfer ports, and the dual port architecture are defined by Fibre Channel standards.

Figure 3 depicts an embodiment in which there are only two computer hosts accessing a particular cluster of Fibre Channel arbitrated memory units. However, it can be extended to include more hosts. The number of hosts that have access to a particular cluster of Fibre Channel arbitrated drives is limited by the number of data transfer ports on the memory storage units. The memory storage units shown in Figure 3 only have two data transfer ports. The number of data transfer ports, however, can be increased. Correspondingly, the number of computer hosts able to directly access the cluster of memory units increases.

Figure 3 depicts an embodiment in which there are only two memory unit clusters connected to each computer host. However, the number of different clusters that can be accessed by a particular computer host can be increased merely by increasing the number of controllers on the computer host.

By extending the symmetric usage of the two ports, it is possible to achieve performance benefits in addition to sharing of information between hosts. When a single port of a dual port drive is active and the other port is in passive standby for fault tolerance, only one host can access a particular disk in a loop at a time. The importance of this condition can be illustrated through the following example. As shown in Figure 1, when the computer host 102 is accessing a first disk drive 106, all of the other disk drives 108, 110, 112 are blocked from being accessed. However, as shown in figure 3, while accessing a first disk drive 302 with a first host 312 through loop 1a,

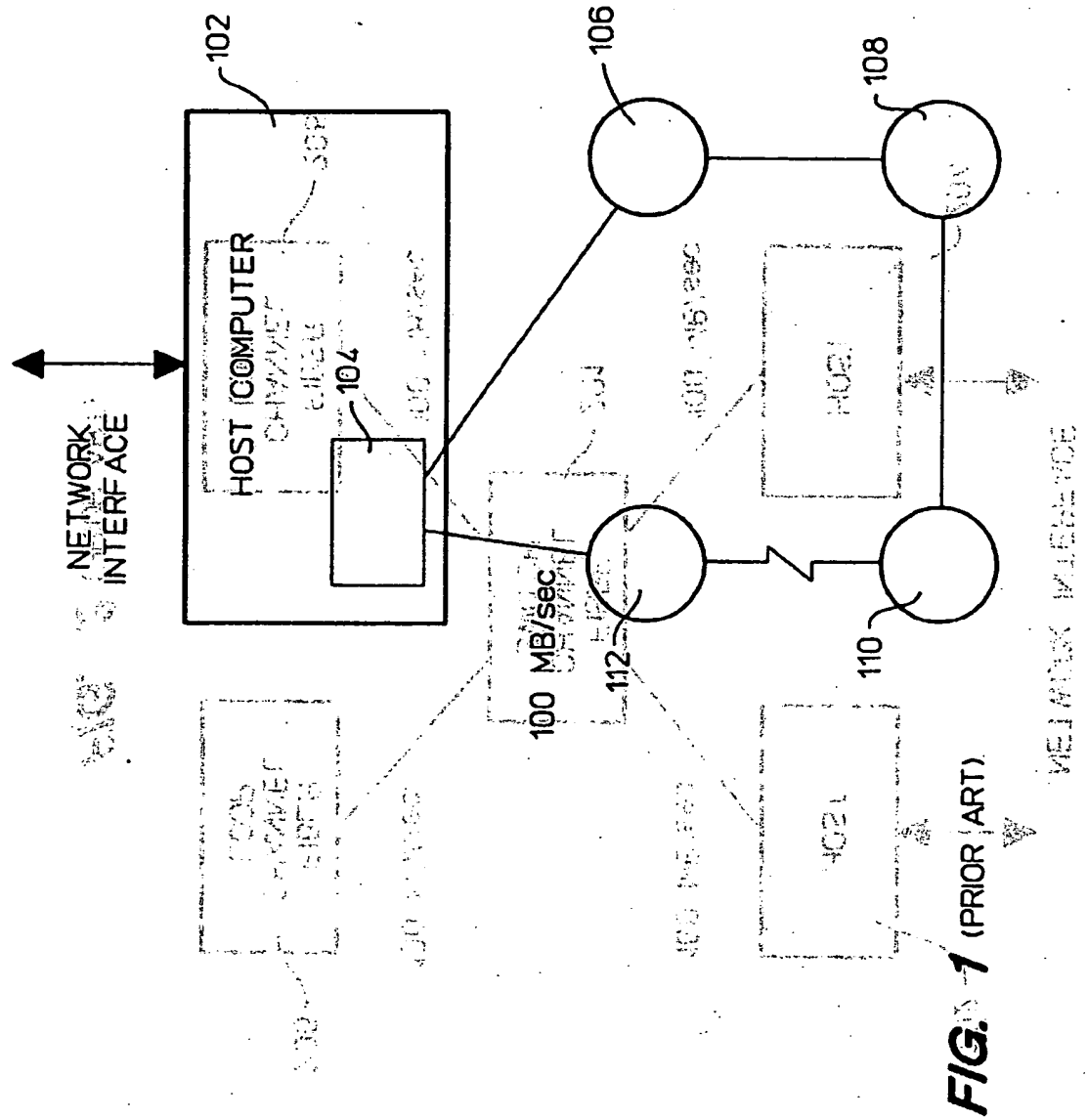
a second disk drive 304 can be accessed by a second host 314 simultaneously through loop 1b.

The disclosures in United States patent application no. 08/494,457, from which this application claims priority, and in the abstract accompanying this application are incorporated herein by reference.

Claims

1. A multi-host information storage system architecture comprising:

- a first memory storage unit cluster (301) including a plurality of memory storage units (302, 304, 306, 308), each memory storage unit including a first data transfer port and a second data transfer port, the first data transfer ports of all the memory storage units being interconnected to form a first loop (1a) and the second data transfer ports of all the memory storage units being interconnected to form a second loop (1b);
- a second memory storage unit cluster (315) including a plurality of memory storage units (316, 318, 320, 322), each memory storage unit including a first data transfer port and a second data transfer port, the first data transfer ports of all the memory storage units being interconnected to form a third loop (2a) and the second data transfer ports of all the memory storage units being interconnected to form a fourth loop (2b);
- a first host (312) including a first controller (310) in communication with the first loop (1a) and a second controller (326) in communication with the third loop (2a); and
- a second host (314) including a first controller (324) in communication with the second loop (1b) and a second controller (328) in communication with the fourth loop (2b).
2. An architecture as in claim 1, wherein:
 - a first data transfer port of each of the memory storage units in the first memory storage unit cluster (301) including an input and an output; the input of the first data port of one of the memory storage units is in communication with an output of the first controller of the first host (312); and
 - the output of the first data port of another one of the memory storage units is in communication with an input of the first controller of the first host (312); and
 - the remaining inputs and outputs of the first data ports are interconnected to define the first loop (1a).
 3. An architecture as in claim 1 or 2, wherein at least one of the loops is a Fibre Channel arbitrated loop.
 4. An architecture as in claim 1, 2 or 3, wherein the transfer ports are interconnected optically.
 5. An architecture as in claim 1, 2 or 3, wherein the transfer ports are interconnected electrically.
 6. An architecture as in any preceding claim, wherein the memory storage units are disk drives.



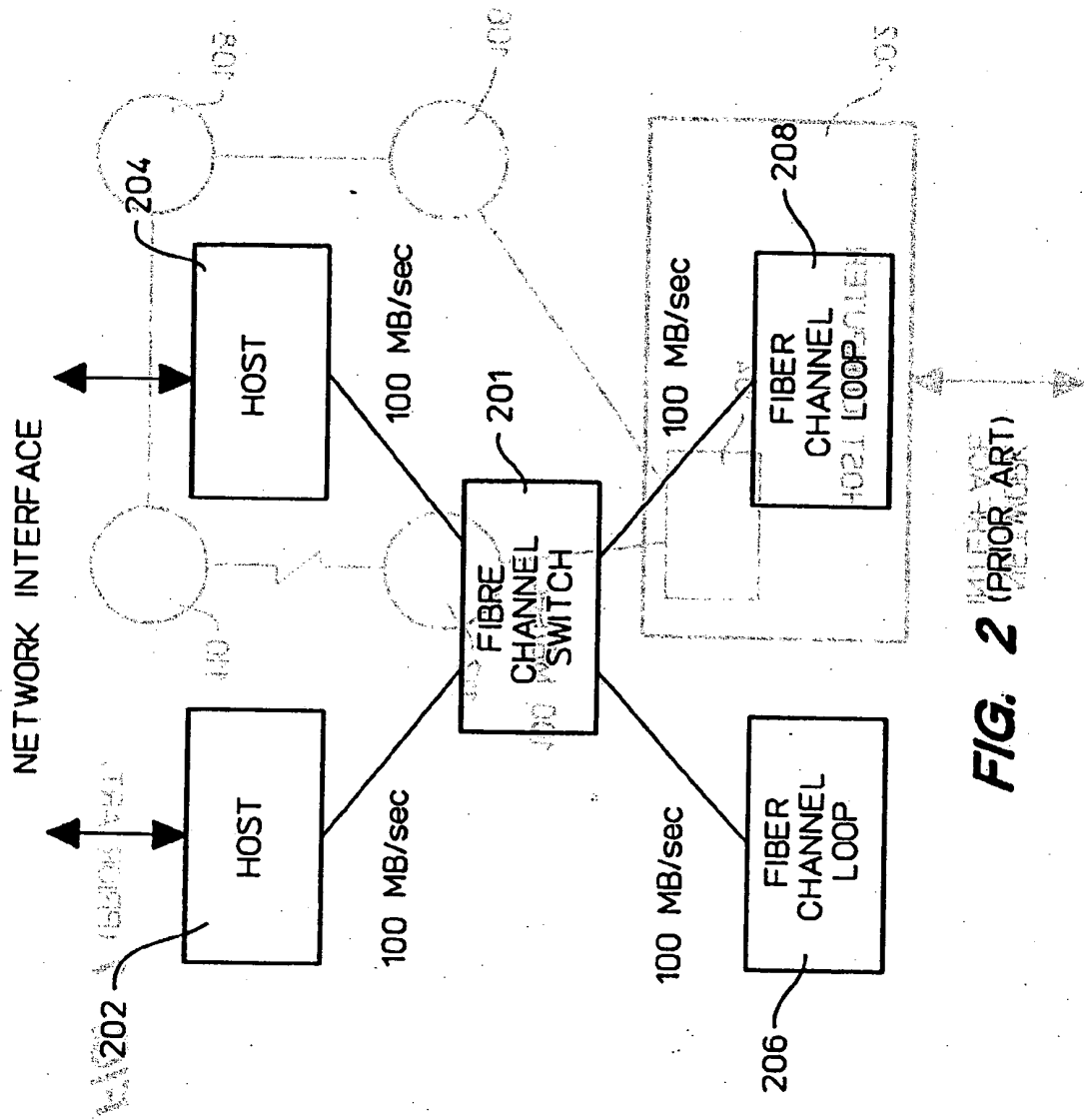


FIG. 2 (PRIOR ART)

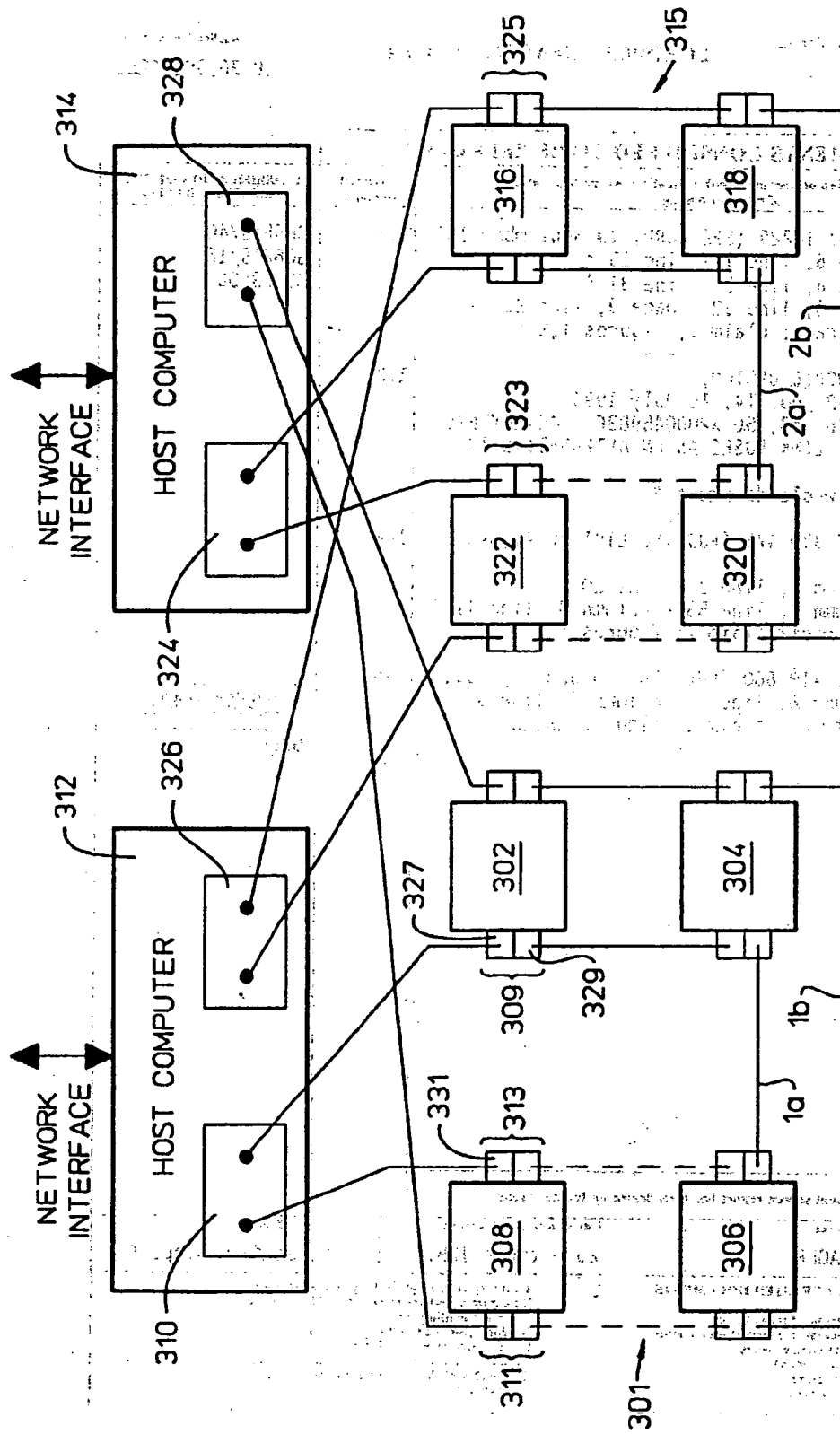


FIG. 3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 30 3626

DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim
Y	WO-A-91 14229 (SF2 CORP) 19 September 1991 * page 6, line 1 - line 23 * * page 8, line 5 - line 31 * * page 8, line 32 - page 9, line 21 * * abstract; claim 1; figures 1,2 *	1,2
Y	ELECTRONIC DESIGN, vol. 42, no. 14, 11 July 1994, page 46, 48, 50 XP000459836 *40-MBYTE/S SERIAL LINK POSES AS AN ALTERNATIVE TO SCSI* * the whole document *	1,2
A	EP-A-0 330 475 (FUJITSU LTD) 30 August 1989 * column 1, line 3 - line 20 * * column 2, line 58 - column 3, line 11 * * abstract; claim 1; figures 1,3 *	1-6
A	EP-A-0 410 860 (BULL SA) 30 January 1991 * column 6, line 1 - column 7, line 3 * * abstract; claim 1; figures 1A,1B *	1-6
The present search report has been drawn up for all claims		
Place of search	Date of completion of the search	Examiner
THE HAGUE	23 October 1996	Nguyen Xuan Hiep, C
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure F : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application I : document cited for other reasons & : member of the same patent family, corresponding document</p>		

EPO FORM 1503 (01.92) (P0401)